

Utah State University

DigitalCommons@USU

All Graduate Theses and Dissertations

Graduate Studies

5-1968

The Relationship of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) to the Stanford-Binet Intelligence Scale

Duane Reeder
Utah State University

Follow this and additional works at: <https://digitalcommons.usu.edu/etd>



Part of the [Psychology Commons](#)

Recommended Citation

Reeder, Duane, "The Relationship of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) to the Stanford-Binet Intelligence Scale" (1968). *All Graduate Theses and Dissertations*. 5650.

<https://digitalcommons.usu.edu/etd/5650>

This Thesis is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



THE RELATIONSHIP OF THE WECHSLER PRESCHOOL
AND PRIMARY SCALE OF INTELLIGENCE
(WPPSI) TO THE STANFORD-BINET
INTELLIGENCE SCALE

by

Duane Reeder

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Psychology

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1968

ACKNOWLEDGMENTS

I would like to thank Dr. Glendon Casto, chairman of my Graduate Committee, for his support, encouragement, and critical review of this thesis. He is also to be thanked for his ideas which led to this thesis and the patience which he extended to me during the time this study was being researched and prepared.

Mrs. Marilyn Black and Mrs. Loree Speuler, the teachers of the Head-Start students, who were used as subjects for this thesis, are to be recognized for their very fine cooperation and consideration during the time when the testing was being done.

Others who assisted materially included Mr. Sam Campenella and Mr. Edward Allen, fellow graduate students in Psychology. Special gratitude is extended to Dr. Worten Allen who was gracious enough to replace Dr. E. Wayne Wright in his absence and Dr. Michael Bertoch who offered his counseling and support.

Finally, to my wife, Melva, for her patience and support in fulfilling this assignment, I extend my love and appreciation.

Duane Reeder

TABLE OF CONTENTS

	Page
INTRODUCTION	1
Problem Statement	2
Objective.	4
REVIEW OF LITERATURE.	5
History of Head-Start.	5
Head-Start Qualifications and Characteristics of Students.	5
Infant and Preschool Tests	6
Hypotheses	11
METHOD OF PROCEDURES.	13
DESCRIPTION OF TESTS USED	15
The Wechsler Preschool and Primary Scale of Intelligence (WPPSI)	15
Comments on New WPPSI Subtests.	17
The Stanford-Binet.	19
RESULTS	22
Hypothesis 1.	22
Hypothesis 2.	23
Hypothesis 3.	24
SUMMARY AND CONCLUSIONS.	27
LITERATURE CITED	29
VITA	31

LIST OF TABLES

Table	Page
1. New test items in WPPSI and number retained from WISC . .	17
2. Correlation table comparing Stanford-Binet and WPPSI scores.	22
3. Relationship between Wechsler's correlations and this study's correlations.	24
4. Percent r's between the scale of the WPPSI with the Stanford-Binet.	25

ABSTRACT

The Relationship of the Wechsler Preschool And Primary Scale of Intelligence (WPPSI) to the Stanford-Binet Intelligence Scale

by

Duane Reeder, Master of Science

Utah State University, 1968

Major Professor: Dr. Glendon Casto

Department: Psychology

Correlational comparisons were made between the Stanford-Binet, Form L-M, and the Wechsler Preschool and Primary Scale of Intelligence using children enrolled in a Head-Start program. The study was concerned with three hypotheses:

1. The correlations found between the I.Q. scores obtained on the WPPSI full scale, verbal, and performance scales and those obtained on the Stanford-Binet using Head-Start children as subjects would be significant at the .01 level.
2. The correlation between the WPPSI and the Stanford-Binet utilizing Head-Start children would not be significantly different from the correlation reported by Wechsler in the WPPSI manual.
3. Scores on the WPPSI verbal scale and the Stanford-Binet would correlate higher than would the WPPSI performance scale scores with the Stanford-Binet.

All correlations run relating to the three hypotheses chosen for this study were found to be significant at the .01 level. The results, therefore, lead to the acceptance of all three hypotheses.

(37 pages)

INTRODUCTION

Because of the Stanford-Binet's long-range stability, usefulness in making predictions, and the mass of research it has generated in the area of intelligence, it has long been a standard procedure to evaluate and criticize the usefulness of new intelligence tests by comparing them with the Stanford-Binet. Anastasi (1961) has gone as far as stating that the Stanford-Binet Intelligent Quotient has almost become synonymous with intelligence itself.

There are, however, two other intelligence tests which are widely used by behavioral scientists. These are the Wechsler Adult Intelligence Scale and the Wechsler Intelligence Scale for Children, commonly known as the WAIS and WISC, respectively.

Since the publication of the WISC in 1949, Wechsler has received numerous requests to develop a downward extension of the instrument. Many have felt a need for an instrument like WISC to test the I.Q. of preschool children (Wechsler, 1966). The lowest age testable with the WISC is five years. Although many preschool and infant scales do exist, they are not being used as extensively as the Stanford-Binet or

the WISC due to defects found in standardization, unsatisfactory reliability, lack of appeal for the children and I.Q. conversions which are of questionable accuracy at this particular age range (Cronbach, 1960).

As a result of the requests made for a downward extension of the WISC, Wechsler and the Psychological Corporation released in December of 1966, a new instrument called the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), which was to be used in measuring the I.Q. of children from age four to six-and-one-half years.

As of yet, little is known concerning the WPPSI's validity and it's reliability using various types of populations or it's ability to withstand the test of time and the numerous criticisms that are constantly being hurled at the infant and preschool instruments that are already in existence. However, the Stanford-Binet has stood the test of time, has been recognized as "the test of intelligence," and is considered to be one of the most useful intelligence instruments in use.

Problem Statement

The Stanford-Binet Intelligence Scale is considered to be one of the leading, if not the leading, instruments for measuring intelligence. No other instrument has been researched and studied more extensively or has accumulated more knowledge about intelligence than has the Stanford-Binet. It has been the test by which many other tests have been judged. It seems therefore to be the best instrument available by which the new Wechsler Preschool and Primary Scale of Intelligence can be evaluated and validated for use as a useful tool in the psychometric field.

The 1960 revision of the Stanford-Binet is a revision of earlier Binet tests and continues to meet Binet's description of intelligence,

"the tendency to take and maintain a definite direction; the capacity to make adaptations for the purpose of attaining a desired end; and the power of auto-criticism," (Cronbach, 1960, p. 160) and to fit his belief that intelligence is general in nature. Binet's test, and all those that have been patterned after it, including the 1960 revision, have been designed to measure only "general intelligence."

Wechsler, in constructing his tests (WAIS, WISC, and WPPSI), has agreed with Binet's idea of general intelligence, but he also recognizes that in certain areas, a persons' ability to function is better than in other areas. With this in mind, Wechsler divided his tests into a number of subtests, and each related to a specific type of item. This has made it possible for the skilled clinician to determine diagnostic scores on the various kinds of behavior which each subtest measures. The subtests on the WPPSI are grouped into two categories. One category gives "verbal I.Q.'s" the other "performance I.Q.'s." These characteristics indicate that the Wechsler tests are perhaps more useful in diagnosis than the Stanford-Binet. However, it is not the purpose of this study to evaluate the diagnostic value of the WPPSI.

The problem that this study is concerned with is that there exists a complete lack of research involving comparisons between the Stanford-Binet and the WPPSI using children who are enrolled in specialized school systems such as HEAD-START. Wechsler (1966) has indicated a need for further comparisons between the WPPSI and other instruments using different types of individuals other than those which were used in his stratified sample. The children from Head-Start, by the very nature of their poverty and lack of adequate socialization, represent a different type of

child and will help fill this need.

Furthermore, as recent as the Head-Start program is, research involving I.Q. comparisons between various preschool tests and the children enrolled in this program is difficult to come by.

Objective

It will be the objective of this study to compare the WPPSI I.Q. scores of Head-Start children with their Stanford-Binet I.Q. scores in an attempt to add additional validating data to the WPPSI.

REVIEW OF LITERATURE

The review of literature will be concerned with the following areas: (1) the history of Project Head-Start; (2) the qualifications for entering Head-Start; and (3) a review of a number of infant and preschool tests and their relationship to the Stanford-Binet. The last section of this chapter will state the hypotheses chosen for this study.

History of Head-Start

On February 19, 1964, Mrs. Lyndon B. Johnson, wife of the President of the United States, met with some 250 women leaders at the White House. the purpose of this meeting was to introduce to these women a program called Project Head Start, of which Mrs. Johnson is honorary chairwoman. This program was to assist the nation's communities in establishing schools for children under the age of six years, who, because of financial and cultural deprivation, were listed on the nation's poverty roles. This program is one of several now in existence which has been established since the Economic Opportunity Act of 1964 was passed by Congress. All of the programs now in existence under the Economic Opportunity Act have been established in an effort to help the poverty stricken people of our nation (Meyer, 1965).

Head-Start Qualifications and Characteristics of Students

The requirement that must be met to qualify for this program is strictly a financial one. For example, a family of four with an annual income of \$3,200 or less would qualify for the program. Those children

who are "culturally deprived" are given first priority. Children are considered to be culturally deprived if they come from broken homes, or if the parents are unemployed, or if the child's family is on welfare.

Some of the noticeable characteristics of the Logan Head-Start children are as follows: (1) many did not know how to use eating utensils (knives, forks, spoons, etc.) and therefore ate with their fingers; (2) a complete lack of discipline existed when the children first entered the program; (3) many of the children used objectionable language (swearing); (4) all of them didn't know how to share with others; (5) many of the children didn't know or recognize the differences in foods (meat, potatoes, vegetables, and fruits) or between colors; (6) some couldn't tell the differences between animals such as cattle and horses; (7) most couldn't dress themselves; (8) some were not toilet trained; and finally, (9) most of the children were extremely aggressive with a few being almost totally repressed and passive in nature.

Infant and Preschool Tests

Today there are a number of infant and preschool tests on the market. A few of them are: (1) the Gesell Developmental Schedules, which measure infants and children from age four weeks to six years; (2) the Cattell Infant Intelligence Scale, which has as its age range two to thirty months; (3) the California First-Year Mental Scale tests from age one year to eighteen months; (4) the Merrill-Palmer Scale which measures I.Q.'s of children from age two to five years; (5) the Goodenough Draw-a-Man Test (DAM) which measures not only intelligence, but personality development as well. It has an age range of one to ten years; (6) the Columbia Mental Maturity Scale, ages three to twelve; and, (7) the Pea-

body Picture Vocabulary Test (PPVT), which is the newest of the infant and preschool tests listed to join the files of tests that are currently (Anastasi, 1961 and Cronbach, 1960) available.

Even though there are numerous tests available which deal with infants and children from the age of a few months to several years, studies (Bayley, 1949 and Pinneau, 1961) have shown that those tests which claim to measure intelligence from a few months to two or three years do not yield I.Q.'s which are stable or valid. The two studies just cited have shown that I.Q.'s can vary by as many as thirty points from the first time the child was tested until one or two years later when the child was retested. Bayley (1955) has gone as far as stating that "none of these efforts has been successful in devising an intelligence scale applicable to children under two years that will predict their later performance" (p.807). However, these same investigators have stated that the stability of I.Q. scores increases rapidly starting with children who are two years old and older. This increase in stability tends to make any predictions that might be made, concerning learning ability as it relates to I.Q. scores, more valid and reliable. The issue of I.Q. stability is the main reason why the Stanford-Binet has as its lowest age two years.

Two studies concerning the use of the Columbia Mental Maturity Scale (CMMS) and its relationship to the Stanford-Binet yield somewhat conflicting results. Cronbach (1960) reports a correlation of about .75 between these two instruments. No other information is given concerning individual studies, but this does lead one to believe that these two instruments do have much in common and tend to measure the same types of

information which lead to determining an I.Q. score. On the other hand, Levinson and Block (1960) report a correlation of .45 between the CMMS and the Stanford-Binet mental age and .39 between the CMMS and the Stanford-Binet I.Q. These two investigators report that the CMMS I.Q.'s were consistently lower than the Stanford-Binet I.Q.'s. Based upon this study one is led to believe that these two instruments do not measure the same characteristics. However, Levinson and Block (1960) used only thirty-nine subjects who ranged in ages for 4.0 to 5.9 years. This small a sample is not an adequate basis for a generalization.

There has been a mass of research comparing the Goodenough Draw-a-Man (DAM) and the Stanford-Binet. In a study conducted by Sundberg (1961) concerning tests which are used most frequently in clinical service, he found that the DAM ranked ninth. This should indicate just how extensively this instrument is used. As a matter of information, Sundberg (1961) also found that the Stanford-Binet ranked fifth and the WISC tenth. Although this study was not concerned with the validity of these various tests, it does give us an indication of which instruments are the most popular.

Thompson and Findley (1963) tested 164 children with both the Stanford-Binet, form L-M, and the Goodenough Draw-a-Man test to first compare the mean I.Q. scores of the Stanford-Binet and the DAM; second, to find the correlation between the Stanford-Binet and the DAM scores; and third, to analyze the difference between the Stanford-Binet and the DAM scores. The results showed that the I.Q.'s on the Stanford-Binet were on the average 4.15 points higher than the I.Q. scores on the DAM. The correlation found between the two instruments was .67. In analyzing the difference

between the I.Q. points, it was found that forty-eight percent of the sample varied more than twelve points between these two tests with the lowest I.Q.'s being obtained on the DAM. This led the authors to question the value of using the DAM for assessing the I.Q.'s of children.

Estes, Curten, DeBurger, and Denny (1961) compared the 1937 form of the Stanford-Binet, the 1960 revision of the Stanford-Binet, the WISC, the Raven and the DAM with each other in an effort to see if the 1960 revision of the Stanford-Binet would differ significantly from the 1937 Stanford-Binet and the other instruments involved. These people found that age was not a factor when comparing the different revisions of the Stanford-Binet and the WISC. They also found that when comparing average groups of students with the various instruments that no significant difference was found in their I.Q. scores. But, they did find a significant difference in I.Q. scores between the instruments involved when testing students who fell into the superior range. The authors found the correlation between the DAM and the 1960 Stanford-Binet to be .43; between the DAM and the 1937 Stanford-Binet .46 to .41; between the DAM and the WISC .43. The Raven correlated .59 with the 1960 Stanford-Binet, .67 to .54 with the 1937 Stanford-Binet and .55 to .91 with an average correlation of .75 with the WISC. Rohrs and Haworth (1962) using mentally retarded children as subjects reported a correlation of .28 between the DAM and the 1960 Stanford-Binet. Thompson and Finley (1963) in their review of the literature report correlations from .36 to .80 between the DAM and the Stanford-Binet. The large range in correlations reported by the various investigators, with respect to these two tests, makes one feel that the DAM is not consistent enough to base much confidence in its

ability to yield I.Q.'s which are stable and valid. If the above assumption is true, then this adds evidence to support Wechsler's statement that there is need for an instrument like the WISC to measure I.Q. scores in preschool children (1960).

In 1959, Dunn developed the Peabody Picture Vocabulary Test to measure the verbal intelligence of the mentally handicapped. Mein (1962), using as subjects eighty patients from two mental hospitals near London, found a correlation of .71 between the vocabulary age on the Peabody and the mental age on the Stanford-Binet.

Milgram and Ozer (1967) using as their subjects two groups of Negro children from impoverished homes in a large Eastern city, who were enrolled in a Head-Start Project, found that both groups of students received I.Q. scores which were significantly below the average (average is 100 I.Q. points on the Stanford-Binet) when they were tested on both the Peabody and the Stanford-Binet. They also found that the I.Q.'s obtained on the Peabody were significantly lower than the I.Q. scores on the Stanford-Binet. The first groups of students were administered the Peabody twice and the Stanford-Binet twice. The correlations found for these two administrations were .60 and .45. In the second group only one administration of the Peabody and the Stanford-Binet was given and the resulting correlation was .44. Again, the correlations reported in this study are not considered high.

Up to this point the question might be asked, what is the reason for the development of a new preschool test when so many are available? The main reason is that most of the preschool and infant tests which do exist do not yield I.Q. scores which are stable enough or valid enough

to make predictions from. According to Schacter and Apgar (1958) the only instrument which exists that can be used in predicting the future performance of a preschool child is the Stanford-Binet. However, the Wechsler Intelligence Scale for Children has been used extensively for making predictions concerning achievement which have been found to be valid (Frandsen and Higginson, 1951). A major problem is that the WISC has as its lowest age limit five years and this is not considered to be preschool age.

The only study which has been reported comparing the WPPSI and the Stanford-Binet is that which is reported by Wechsler in the WPPSI manual (1966). He reports a correlation of .75 between these two instruments using a stratified sample of the United States population. He also reports a correlation of .58 between the WPPSI and the Peabody Picture-Vocabulary Test and .64 between the WPPSI and the Picture Test of Intelligence.

In conclusion, it may be stated that there does exist a great amount of interest in the area of preschool testing and that most of the tests that have been available do not yield I.Q.'s which are stable and upon which valid predictions can be made.

Hypotheses

The hypotheses chosen for study in this thesis are:

1. The correlations found between the I.Q. scores obtained by the WPPSI full scale, verbal, and performance scales and those obtained on the Stanford-Binet using Head-Start children as subjects will be significant at the .01 level.
2. The correlation between the WPPSI and the Stanford-Binet

utilizing Head-Start children will not be significantly different from the correlation reported by Wechsler in the WPPSI manual (1966).

3. Scores on the WPPSI verbal scale and the Stanford-Binet will correlate higher than will the WPPSI performance scale scores with the Stanford-Binet.

METHOD OF PROCEDURE

The students who were enrolled in a Head-Start Program at the Wilson elementary school, Logan, Utah, during the summer months of the year 1967, were used for the research. A total of twenty-five children were enrolled and of that total, twenty-two children were used as subjects for this study. There were three students who were not included in this study due to the fact that they were moving to inaccessible areas and would not be available for follow-up studies being used in another Master's thesis. The ages of the subjects ranged from four years one month to five years eight months. Both the 1960 revision, Form L-M, of the Stanford-Binet Intelligence Scale and the WPPSI were administered to each pupil. Both instruments were administered by trained personnel and were given according to the directions in the testing manuals. All subjects were administered both tests within sixty days of each other.

In an effort to test the hypotheses stated previously, these two tests were correlated using the Pearson-product Moment Correlation. A correlation was run between the full scale I.Q. scores of the WPPSI and the Stanford-Binet. A correlation was also run between the performance I.Q. scores on the WPPSI and the Stanford-Binet. Another correlation was run using the verbal I.Q. scores on the Stanford-Binet. In order to compare correlations found by Wechsler (1966) between the full scale WPPSI and the Stanford-Binet with the correlation found in this study, Fisher's Z or Transformation statistic was used. Each correlation run

was followed by a t-test, $t = r \frac{N-2}{1-r^2}$, to test the level of significance of each of the correlations found.

In addition to the correlations run between the WPPSI and the Stanford-Binet, comparisons were also made between the WPPSI full scale I.Q. scores and the performance and verbal I.Q. scores. A correlation was also run between the performance and verbal I.Q. scores using the Pearson-Product Moment Correlation.

DESCRIPTION OF TESTS USED

The Wechsler Preschool and Primary Scale of Intelligence (WPPSI)

The WPPSI, published December, 1966, is a test designed to obtain intelligence scores of children in the form of deviation I.Q.'s. The age range is from four to six-and-one-half years. Due to the various interpretations and ambiguities which exists when dealing with the concept of mental age (MA), Wechsler has discarded it in favor of the deviation I.Q. (Wechsler, 1966). Each item in the WPPSI is credited a specified number of points, the points earned being added to determine the raw score on that particular test. The total number of points earned is converted, using norm tables, into a deviation I.Q. The WPPSI is divided into eleven subtests, each containing one type of task. Breaking the WPPSI into separate subtests was done because it has been diagnostically useful in the past on both the WISC and WAIS. The various subtests are grouped into two series, one which yields a "verbal I.Q.," and the other a "performance I.Q." When both the verbal and performance I.Q.'s are totaled they yield a "full scale I.Q."

The WPPSI verbal scale includes tests of Information, Vocabulary, Arithmetic, Similarities, Comprehension, and Sentences, which is considered a supplementary test. The performance scale includes Animal House, Picture Completion, Mazes, Geometric Design, and Block Design.

Unlike the WISC, both verbal and performance tests are administered intermixed. This change in administration from the WISC was made in an

effort to be better able to insure interest and cooperation in the young child.

"Of the eleven tests which constitute WPPSI, eight (Information, Vocabulary, Arithmetic, Similarities, Comprehension, Picture Completion, Mazes, and Block Design) are tests reincorporated from the WISC." (Wechsler, 1966, p.7). The remaining three tests (Sentences, Animal House, and Geometric Design) are new and take the place of the Digit Span, Picture Arrangement, Object Assembly and Coding, which could not be used in the WPPSI for the following reasons:

Digit Span was discarded primarily for its limited score spread. Coding was omitted because of the uncertain role of the motor factor and the influence which previous exposure to the use of pen and pencil had on the performance of many children. Picture Arrangement was omitted from the WPPSI because the test proved too difficult for 4-and-5-year-olds; only a very small percentage of children at this age level were able to grasp the intent of the test. Object Assembly had to be dropped (after being administered in preliminary tryout to several hundred subjects) because of its low test reliability; in most other respects the test was satisfactory. (Wechsler, 1966, p.8).

Table 1 gives an account of new test items introduced in the WPPSI and the number retained from the WISC.

Table 1. New test items in the WPPSI and number retained from WISC.

Test	Number of Items		
	<u>New</u>	<u>From WISC</u>	<u>Total</u>
Verbal:			
Information	11	12	23
Vocabulary	8	14	22
Arithmetic	14	6	20
Similarities	9	7	16
Comprehension	9	6	15
Sentences (Supplementary test)	13	0	13
Performance:			
Animal House	20	0	20
Picture Completion	11	12	23
Mazes	3	7	10
Geometric Design	10	0	10
Block Design	4	3	7

Wechsler (1966) did modify somewhat those tests which were carried over from the WISC by adding easier items and eliminating some of the more difficult ones.

Comments on New WPPSI Subtests

The Animal House subtest resembles the Coding test on the WISC in that it requires the child to make an association between signs and symbols. Wechsler (1966) states that this test measures such things as memory, attention span, goal awareness, the ability of the child to concentrate or in general terms, this test measures the child's ability to learn.

The test of Geometric Design was added to the WPPSI because it had

been shown previously that a child's ability to reproduce geometric designs correlates well with other measures of intelligence. This test also gets away from some of the limitations that are associated with verbal tests. Visual-motor coordination and perception are measured by this test.

Sentences take the place of the Digit Span test used in the WISC. The major difference is that a child is given some credit for partial recall in the Sentence test where he is not in the Digit Span test on the WISC.

The WPPSI was standardized on a sample of 1,200 children, 100 boys and 100 girls in each of the six age groups from age four to six-and-one-half at half year intervals. This sample population was gathered from four geographical regions in the United States. The areas represented were: (1) Northeast; (2) North Central; (3) South; and (4) West. Eight occupational categories were established for the purposes of stratifying the sample. The professional areas represented are: (1) Professional, technical, and kindred workers; (2) Managers, officials, and proprietors (except farm); (3) Clerical, sales, and kindred workers; (4) Craftsmen, foremen, and kindred workers; (5) Operatives and kindred workers; (6) Service workers, including private household; (7) Farmers and farm managers; and, (8) Laborers, including farm laborers. These eight groups were a condensation of the 1960 Census. Of the 1,200 children used in the standardization, 85.8% were white and 14.2% nonwhite, 68% were from urban areas while 32.0% were from rural areas.

Each child tested with the WPPSI is assigned an I.Q. score. The I.Q.'s are based on scaled scores which come from each age group. As

has been mentioned previously in this paper the I.Q.'s obtained on the WPPSI are deviation I.Q.'s in that the I.Q. score indicates how much a child deviates above or below the average score of individuals of his own age group. The I.Q. of 100 and a standard deviation of 15 was predetermined. In terms of percentiles the highest 2.2 percent will have I.Q.'s of 130 or above and 2.2 percent of 69 or below. The middle 50 percent of children in each age group will have I.Q.'s of 90-109.

(Wechsler, 1966, p. 20)

The Stanford-Binet

The 1960 revision of the Stanford-Binet, like the WPPSI, is an individual intelligence test. The 1960 revision is the result of incorporating into a single scale the best subtests from forms L and M of the 1937 revision. So in essence, the 1960 scale is the L-M form.

The major changes in the 1960 revision of the Stanford-Binet are of two types; content and structural. The changes in content material were in the direction of eliminating the subtests or items which have been less satisfactory, eliminating those items which duplicate each other, relocating those items that have proven to be satisfactory, and rescored where it became necessary. Any changes made with those items which were retained from the 1937 revision were held to a minimum.

The changes which are considered to be the most radical have been in the area of correcting structural inadequacies which have existed in the 1937 revision. In the 1937 Stanford-Binet the mean I.Q.'s at various stages were somewhat above 100 (Terman and Merrill, 1937). This was corrected in the 1960 Stanford-Binet by adjusting the mean I.Q. of each chronological age which exceeded 100 so that it became 100.

The two major structural changes which were made were: (1) changing the conventional I.Q. used in the 1937 revision ($I.Q. = \frac{MA}{CA} \times 100$) into deviation I.Q.'s and (2) extended the I.Q. tables to include ages seven-teen and eighteen. These changes were made to help eliminate some of the inadequacies found to exist using the conventional method of establishing I.Q. scores and because retest findings have shown that mental growth extends past age sixteen (Bradway, Thompson, and Crave, 1958). The deviation I.Q. being used in the 1960 scale is basically a standard score with a mean of 100 and a standard deviation of 16.

The 1960 revision of the Stanford-Binet used the stratified sampling method of standardization. Children and students from ages two-and-a-half to eighteen years of age from six different states (California, Minnesota, Iowa, New York, Massachusetts, and New Jersey) were used as subjects. A total of 4,498 subjects participated in the standardization. They were not divided proportionally among the various age groups (Terman and Merrill, 1960).

The stratified samples were chosen by grouping the subject's fathers into occupational categories and in the case of the fifteen-year-olds, by grade placement. Six occupational classifications were determined using the 1950 Census figures. The six groupings made were: (1) Professional and technical workers, 8.2 percent; (2) Managers, officials, proprietors, farm managers and farm owners, 21.2 percent; (3) Clerical and sales workers, 12.5 percent; (4) Craftsmen, foremen, and operatives, 40.2 percent; (5) Private household and service workers, 6.6 percent, and, (6) Laborers-farm and non-farm, 11.2 percent (Terman and Merrill, 1960).

When administering the test, the examiner gives those tests which

are somewhat below the child's expected ability. Doing this establishes the basal age of the child or the level at which all tests can be passed by the child. Once the basal age is established, tests of each higher level are given. This procedure continues until the child fails to pass all the tests at some level.

The Stanford-Binet includes a variety of tasks for both verbal and non-verbal performances. Tests measure anything from simple memory recall to complex reasoning ability, from recognizing familiar situations to answers which have to be learned and problems which call for the ability to adapt to novel situations.

RESULTS

In this chapter, data is presented relating to the hypotheses chosen for this study.

Hypothesis 1: The correlations found between the I.Q. score obtained on the WPPSI full scale, verbal, and performance scales and those obtained on the Stanford-Binet, using Head-Start children as subjects will be significant at the .01 level.

The Correlations found between the two intelligence instruments, which were used in this study, are presented in Table 2. Also included in Table 2 are the levels of significance which were attained by the correlations using the t-test formula, $t = r \sqrt{\frac{N-2}{1-r^2}}$.

Table 2. Correlation table comparing Stanford-Binet and WPPSI scores.

	S-B	WPPSI-F	WPPSI-V	WPPSI-P	Level of Sign.
WPPSI-F	.78				.01
WPPSI-V	.80	.95			.01
WPPSI-P	.64	.91			.01
Mean	107.36	99.82	98.60	101.32	
SD	15.15	13.39	12.54	13.59	

As one can see, all of the correlations reported in Table 2 are significant at the .01 level. Therefore, Hypothesis 1 is accepted.

With all correlations being significant at better than the one percent level, it can be said that in 99 out of 100 cases the correlations will not be due to chance fluctuations, but will be due to a true relationship between the two variables.

The correlation of .78 between the WPPSI full scale I.Q.'s and the Stanford-Binet I.Q.'s indicates that in this study's population, and in populations which are similar, one may predict, with better than 60 percent accuracy, that a person receiving an I.Q. score on the Stanford-Binet, Form L-M, will also receive a similar I.Q. score on the WPPSI full scale.

Hypothesis 2: The correlations between the WPPSI scales and the Stanford - Binet will not be significantly different from the correlations reported by Wechsler (1966) in the WPPSI manual.

In order to test this hypothesis, a comparison was made between the correlations found by Wechsler (1966) and those reported in this study between the Stanford-Binet and the WPPSI full scale, verbal, and performance scales, using Fisher's Z or Transformation statistic of

$$(Z = \frac{Zr_1 - Zr_2}{\sqrt{1/(N-3) + 1/(N-3)}})$$

Table 3. Relationship between Wechsler's (1966) correlations and this study's correlations.

Tests	r	Z	Level of Confidence
WPPSI-F with S-B (Wechsler)	.75	.368	.01
WPPSI-F with S-B (this study)	.78		
WPPSI-V with S-B (Wechsler)	.76	.412	.01
WPPSI-V with S-B (This study)	.80		
WPPSI-P with S-B (Wechsler)	.56		
WPPSI-P with S-B (This study)	.64	.500	.01

Looking at the results in Table 3, one notices that there are no significant differences at the .01 level of confidence (1.96) between the correlations reported by Wechsler (1966) and those found in this study between the Stanford-Binet and the three scales of the WPPSI. Therefore, Hypothesis 2 is accepted.

Hypothesis 3: Scores on the WPPSI verbal scale and the Stanford-Binet will correlate higher than will the WPPSI performance scale scores with the Stanford-Binet.

Again, looking at Tables 2 and 4, one notes that the correlation between the Stanford-Binet and the WPPSI verbal score is .80 while the correlation between the Stanford-Binet and the performance scale on the WPPSI is .64. Both correlations are significant at the .01 level.

These results confirm Hypothesis 3.

In addition to providing evidence which support the three hypotheses, computations were made by squaring each correlation to give some indication of the present association that exists between the Stanford-Binet and various WPPSI scales. Table 4 provides this information.

Table 4. Percent r 's between the scale of the WPPSI with the Stanford-Binet.

Variables	r	% dependent on common factors	% dependent on different factors
S-B and WPPSI-F	.78	61	39
S-B and WPPSI-V	.80	64	36
S-B and WPPSI-P	.64	41	59
WPPSI-F and WPPSI-V	.95	90	10
WPPSI-F and WPPSI-P	.91	83	17
WPPSI-V and WPPSI-P	.71	50	50

The results in Table 4 show that the highest correlation exists between the WPPSI verbal scale and the WPPSI full scale. The correlation of .95 between these two scales indicates that 90 percent of the time these two scales are measuring common factors and that in 10 percent of the time they are measuring factors which are not common to both scales.

It will also be noted, that a high correlation, .87, exists between the WPPSI full scale and the performance scale showing that 83 percent of the time these two scales measure common factors while only 17 percent of

the time the variance between these two scales can be attributed to factors which are not common to both scales. These results indicate that both the WPPSI verbal and the WPPSI performance scales measure factors which are common to the full scale WPPSI. On the other hand, one notices that in only 50 percent of the time that WPPSI verbal and performance scales measure factors which are common to both scales. This might indicate that the verbal and performance scales get at different aspects of intelligence even though both scales contribute highly to the full scale WPPSI.

The results in Table 4 also show that the Stanford-Binet seems to measure the same factors that are measured by the WPPSI verbal scale to a higher degree (.80) than it does with the performance scale (.71). This tends to indicate that the Stanford-Binet is composed of activities which are more closely related to the activities in the WPPSI verbal scale than it is to the activities in the WPPSI performance scale. This again, adds support to Hypothesis 3.

SUMMARY AND CONCLUSIONS

Since December, 1966, a new intelligence scale has been on the market for use by the behavioral scientist. It has been the purpose of this study to compare this new test, the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) with an intelligence test which has been recognized as "the test of intelligence," that being the Stanford-Binet, Form L-M.

The subjects who participated in this study were twenty-two students enrolled in a Head-Start Program in Logan, Utah. The subjects age ranged from four years one month to five years eight months. All subjects were given both the WPPSI and the Stanford-Binet. Correlations were run between the various scales on the WPPSI (full, verbal, and performance) and the Stanford-Binet. All correlations were found to be significant at or beyond the .01 level (Table 2). The correlations reported in this study show that a high degree of relationship (.78) exists between the Stanford-Binet and the WPPSI full scale I.Q. A higher relationship was found to exist between the Stanford-Binet and the WPPSI verbal scale (.80) with the lowest relationship existing between the Stanford-Binet and the WPPSI performance scale (.64). The results tend to support the general impression that the Stanford-Binet is loaded quite heavily with items that are academically oriented.

The three hypotheses which were chosen for this study have been accepted due to the results which have been found and the results presented

add further validating data to the usefulness of the WPPSI. However, though all hypotheses have been accepted and additional validating data have been presented, caution must be taken in making generalizations concerning the findings of this study because the population used was small in number ($N=22$) and the population is not considered to be a representative sample of all the Head-Start programs throughout the nation. Before the results of this study can be assumed to be representative of the degree of relationship which exists between the two instruments used in this study, more research similar to that which has been conducted in this study should be conducted in other Head-Start centers throughout the nation.

LITERATURE CITED

- Anastasi, Anne. *Psychological Testing*. New York, N.Y., The Macmillian Company, 1961.
- Bayley, Nancy. Consistency and variability in the growth of intelligence from birth to eighteen. *J. genet. Psychol.*, 1949, 75, 1965-196.
- Bayley, Nancy. On the growth of intelligence. *Amer. Psychol.*, 1955, 10, 805-818.
- Bradway, K.P., Thompson, C.W., and Cravens, R.S. Preschool I.Q.'s after Twenty-five Years. *J. educ. Psychol.*, 1958, 49, 278-281.
- Cronbach, Lee J. *Essentials of Psychological Testing*. N.Y., Harper and Row, 1959.
- Downie, N.M., and Heath, R.W. *Basic Statistical Methods*. New York, Harper and Row, 1959.
- Estes, Betsy Worth; Curten, Mary Ellen; DeBurger, Robert, A., and Denny, Charlotte. Relationships between 1960 Stanford-Binet, 1937 Stanford-Binet, WISC, Raven, and Draw-a-Man. *J. consult. Psychol.*, 1961, 25 (5), 388-391.
- Ferguson, George A. *Statistical Analysis in Psychology and Education*, New York, McGraw-Hill Book Company, 1966.
- Frandsen, A.M., and Higginson, J.B. The Stanford-Binet and the Wechsler Intelligence Scale for Children. *H. consult. Psychol.*, 1951, 15, 236-238.
- Levinson, Boris M., and Block, Zelik. (Yeshiva U.) Research note on Columbia Mental Maturity Scale (CMMS) and Revised Stanford-Binet (1) in a preschool population. *J. clin. Psychol.*, 1960, 16, 158-159.
- Mein, R. Use of the Peabody Picture Vocabulary Test with severely sub-normal patients. *Amer. J. ment. Defic.*, 1962, 67(2), 269-273.
- Meyer, Alberta L. Project Head-Start. *Childh. Ed.*, 1965, (Ap), 41, 431-432.
- Milgram, Norman A., and Ozer, Mark N. (Catholic U. of America) Peabody Picture Vocabulary Test scores of preschool children. *Psychol. Reports*, 1967, 20, 779-784.

- Pinneau, Samuel R. Changes in intelligence quotient: Infancy to maturity. Boston, Mass, Houghton Mifflin, 1961.
- Rohrs, F.W., and Haworth, Mary R. The 1960 Stanford-Binet, WISC, and Goodenough Tests with mentally retarded children. Amer. J. ment. Defic., 1962, 66(6), 853-859.
- Schacter, Frances Fuchs, and Apgar, Virginia. (Columbia U.) Comparison of preschool Stanford-Binet and school-age WISC I.Q.'s. J. educ. Psychol., 1958, 49, 320-323.
- Sundberg, Norman D. The practice of psychological testing in clinical service in the United States. Amer. Psychol., 1961, 16, 79-83.
- Terman, L.M., and Merrill, M.A. Measuring Intelligence. Boston, Houghton Mifflin, 1937.
- Terman, L.M., and Merrill, M.A. Stanford-Binet Intelligence Scale, 3rd revision. Boston, Houghton Mifflin, 1960.
- Thompson, Hack M., and Finley, Carmen J. The relationship between the Goodenough Draw-a-Man Test and the Stanford-Binet Form L-M in children referred for school guidance services. Calif. J. educ. Res., 1963, 14(1), 19-22.
- Wechsler, David. Wechsler Preschool and Primary Scale of Intelligence, New York, N.Y., the Psychological Corp., 1966.

VITA

Duane Harold Reeder

Candidate for the Degree of

Master of Science

Thesis: The Relationship of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) to the Stanford-Binet Intelligence Scale

Major Field: Counseling Psychology

Biographical Information:

Personal Data: Born at Brigham City, Utah, March 24, 1941, son of Harold and Prudence Wood Pett; sir name changed to Reeder by court order March 1966; married Melva Rae Westley June 11, 1959; three children--Duane Allen, Ricky Jay, and Shawn Warren.

Education: Attended elementary school in Brigham City, Utah; graduated from Box Elder High in 1959; received the Bachelor of Science degree from Utah State University, with a major in Psychology and a minor in History in 1966; did graduate work and completed requirements for Master of Science degree, specializing in counseling psychology, at Utah State University in 1968.

Professional Experience: 1967 to present, supervising instructor and psychological counselor, Vocational Improvement Program, Logan, Utah.